DOCUMENT RESUME

ED 078 868 LI 004 403

AUTHOR Skelton, Barbara; And Others
TITLE The Use of Citation Linkages and Networks for Information Retrieval in the Social Sciences.
INSTITUTION Bath Univ. of Technology (England). Univ. Library.
PUB DATE Mar 73
NOTE 66p.;(136 References); Design of Information Systems in the Social Sciences (DISISS) Working Paper No. 6

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Bibliographic Citations; Information Networks; *Information Retrieval; *Information Systems; Research; *Social Sciences

ABSTRACT

The Design of Information Systems in the Social Sciences (DISISS) is a research project with the objective of carrying out research necessary for the effective design of information systems in the social sciences, whether by the creation of new systems or the modification of existing systems. This working paper explores the methods available for the grouping and linking of journal articles by citations, and the use of these methods as a means of information retrieval. The study is not only speculative but exploratory, as the data reported was derived from pilot studies conducted on a very small scale. If the concepts expressed here are considered valid and useful, large-scale tests need to be carried out, with machine-readable files. Such tests will not be possible during the DISISS project, but they could well form part of further research. Not all of the uses of citation data concern retrieval, and not all retrieval by means of citation data involves networks. This paper mentions some of these other uses for the sake of completeness. (Related reports are LI004401 and 004402.) (Author/SJ)

ED 078868

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

PERMISSION TO REPRODUCE THIS COPYRIGHTED MATERIAL HAS BEEN GRANTED BY
Bath University Library
TO ERIC AND ORGANIZATIONS OPERATING UNDER AGREEMENTS WITH THE NATIONAL INSTITUTE OF EDUCATION. FURTHER REPRODUCTION OUTSIDE THE ERIC SYSTEM REQUIRES PERMISSION OF THE COPYRIGHT OWNER

Bath University Library

Design of Information Systems in the Social Sciences

Working Paper No. 6

THE USE OF CITATION LINKAGES AND NETWORKS FOR INFORMATION RETRIEVAL IN THE SOCIAL SCIENCES

March 1973

LI 004 403

FILMED FROM BEST AVAILABLE COPY

CONTENTS

		Page
	PREFACE	
1.0	INTRODUCTION	1
2.0	CITATION NETWORKS	4
2.1	Ways of analysing citation data to show relationships and to develop networks	4
2.2	Procedures for retrieval of information using citation data.	8
2.2.1	CITED search	8
2.2.2	SOURCE search	9
2.2.3	DUAL search	9
2.2.4	Differences in the procedures for a CITED and SOURCE search	9
2.2.5	Relevance of articles comprising the citation network	10
2.2.6	The starting reference	11
2.2.7	Preliminary considerations of the data requirements for construction of citation networks by machine.	11
3.0	APPLICATIONS	13
3.1	Packages of information	13
3.2	Current awareness services	14
3.3.	Computer retrieval techniques	14
3.4	Review series	15
3.5	Foreign language translations	15
3.6	Teaching aids	15
3.7	Aids to research in information science	16
4.0	PILOT STUDIES	17
4.1	Studies to identify groups of articles with keywords in the title	17
4.1.1	Short-term memory - a DUAL search	17
4.1.2	Deviant behaviour - a SOURCE search	19
4.2	Studies to identify articles containing new concepts, new ideas and new terms	20
4.3	Pilot Studies: Problems and conclusions	25

Page

5.0 FUTURE WORK 29

REFERENCES

Appendix A. References - short-term memory citation network

Appendix B. References - deviant behaviour citation network

Appendix C. List of DISISS working papers

LIST OF FIGURES

Figure 1.	One-step citation analysis
Figure 2.	A citation network
Figure 3.	Flowchart of procedures for the construction of citation networks
Figure 4a.	Network construction by a CITED search
Figure 4b.	Network construction by a SOURCE search
Figure 5.	'Short-term memory' citation network
Figure 6.	'Deviant behaviour' citation network
Figure 7.	'T-group' citation network
Figure 8.	'Synectics' citation network
Figure 9.	'Generation gap' citation network
Figure 9a.	1969, 1970 and 1971 citations to Mead, M. 1928
Figure 9b.	'Generation gap' citation network extended

PREFACE

DISISS (Design of Information Systems in the Social Sciences) is a research project based at the University of Bath. The objective of the project is to carry out research necessary for the effective design of information systems in the social sciences, whether by the creation of new systems or the modification of existing systems. The project, which is funded by OSTI, commenced in January 1971.

Work on other parts of the project is being reported in a series of working papers which are listed in Appendix C. These, together with an outline of work carried out in 1971 and 1972, can be obtained from the Library, University of Bath, Claverton Down, Bath, BA2 7AY.

The present working paper explores the methods available for the grouping and linking of journal articles by citations, and the use of these methods as a means of information retrieval. The study is not only speculative but exploratory, as the data reported was derived from pilot studies conducted on a very small scale. If the concepts expressed here are considered valid and useful, large-scale tests need to be carried out, with machine-readable files. Such tests will not be possible during the DISISS project, but they could well form part of further research.

Because, so far as we know, the approach put forward here is new, comments on this paper would be especially welcome. Although the data used relates to the social sciences the method is of course applicable in any discipline.

Not all of the uses of citation data concern retrieval, and not all retrieval by means of citation data involves networks. This paper mentions some of these other uses for the sake of completeness.

The outline of this paper was prepared by Michael Brittain, but Barbara Skelton was largely responsible for the drafting. Drafts were read, and numerous contributions made, by Maurice Line, Stephen Roberts and Robert Bradshaw.

1.0 INTRODUCTION

This working paper explores the potential uses of citation analyses for the retrieval of information and in particular the use of citation networks to identify groups of journal articles.

Citation data has been analysed since the 1920s as a guide in the selection of journals for library collections, and to a lesser extent in determining the coverage of secondary services. These uses will probably increase as optimisation of information systems becomes more necessary. Users of information, however, are generally concerned much more with retrieving relevant journal articles and are not directly concerned with the selection policies, whether of libraries or secondary services.

In its simplest form, a citation analysis involves the collection and counting of citations in a limited number of journals (referred to as 'source items'); this results in a list of cited journal titles, or authors, according to their frequency of citation. Lists of cited journal articles have rarely been produced, because the small samples of source journals typically used yield very few articles that are cited frequently. The relationship between source items and cited items[1] is not usually explored in studies of this kind.

By analysing the relationship between source items and cited items, it is possible to derive a structure of linkages, referred to in this paper as a citation network. A citation network reveals the inter-relationships of articles; the frequency of citation of each article is shown and also the actual source articles that make the citations.

Considerable work has been done on the nature, value and use of citations. A recent review of the literature by Hall (1970) indicates that the majority of the work has been concerned with citation counts. A well-known work in this area is that of Brown (1956), who reported lists of frequently cited journals in mathematics, physics,

---

1 A source item is the source from which bibliographic citations are taken; it can be a journal article, monograph or any other work containing a list of references.
A cited item is a work cited (referred to) by the author of an article or monograph (the 'source item').

chemistry, geology, physiology, botany, zoology, and entomology. These lists have from time to time been used as a guide to the selection of journals by libraries. Basically, the same principles are involved in using a frequency list of monographs, although monographs have received much less attention, largely because the frequency with which most of them are cited is relatively low. Recently more refined measures of journal citations have been attempted. For example, the number of citations received by a journal can be expressed as a ratio of the average number of articles it publishes each year (Garfield 1971); Garfield has called this the 'impact factor'. A further recent use of citation data has been the application of clustering techniques (Price and Schiminovich 1968). An important part of DISISS research is the clustering of journal titles to produce groups of titles related by their citations. This work will be reported in a further working paper.

Little work has been carried out with journal articles (as opposed to journal titles). Kessler (1963) analysed the relationship between articles by grouping together articles with citations in common; he calls this 'bibliographical coupling'. Garfield (1963 and 1970), Garfield, Sher and Torpie (1964) and Price (1965) have used citation data to trace the structure of knowledge and the flow of information. All of these studies have necessitated the construction of citation networks in one form or another, but they have been mainly concerned with the interpretation of the network rather than with the method of its construction.

There have been a few evaluative studies on the use of citation indexes for information retrieval. The best known citation index is Science Citation Index (SCI), which has been used as a data base for the construction of networks described in this paper. Waldhart (1964) compared the compilation of a bibliography on lasers from SCI and from five conventional indexing and abstracting services, concluding that SCI produced more unique references than the other services. Spencer (1967) conducted a similar test whilst compiling a bibliography on the drug thalidomide; again, SCI produced more unique references. Martyn (1965) conducted a small test on SCI tracing the subject of gallium as a semi-conductor; he found that for every two relevant articles he retrieved five irrelevant articles. The 'noise' in the system is therefore quite high, but Martyn concluded that if he had searched using article titles, rather than all the citations from each article, the number of irrelevant references might have been well reduced. These studies examine

specifically the use of citation indexes, not the use of citation networks, with which this paper is mainly concerned; but evaluative studies on the use of networks cannot of course be carried out until networks have been constructed and tested.

The main objective of this paper is to examine the method and procedures for the construction of citation networks in order to identify groups of articles. Several pilot studies were undertaken to examine first the method for construction of citation networks, and secondly to evaluate the effectiveness of generating groups of articles. The networks in these studies were compiled by hand. If articles are to be retrieved on a large scale, the methods must be suitable for machine handling. This working paper does not attempt to cover the details of the file structure for machine handling, although some implications for machine requirements are mentioned. The methods and procedures described in this paper are based on 'hand-drawn' networks; once the methods have been explored, large-scale tests can then be undertaken. However, these tests cannot be carried out during the present DISISS project.

The working paper is arranged in the following manner. Section 2 considers the analysis of data for citation networks and describes methods for construction. Three methods have been developed; they are described in this paper as CITED search, SOURCE search and DUAL search. Section 3 discusses briefly some applications of citation networks for information services. The networking technique allows great flexibility for the retrieval of information, and in particular DISISS has considered the packaging of information based on citation networks. The pilot studies are reported in section 4. Two groups of studies were conducted; i) those to identify groups of articles within a subject, and ii) those to identify groups of articles that deal with new concepts and trends. The latter was felt to be important, for any information system in the social sciences must deal adequately with the 'soft' terminology. Finally, in section 5, brief details are given of further work DISISS hopes to carry out on citation networks.

2.0 CITATION NETWORKS

2.1 Ways of analysing citation data to show relationships and to develop networks.

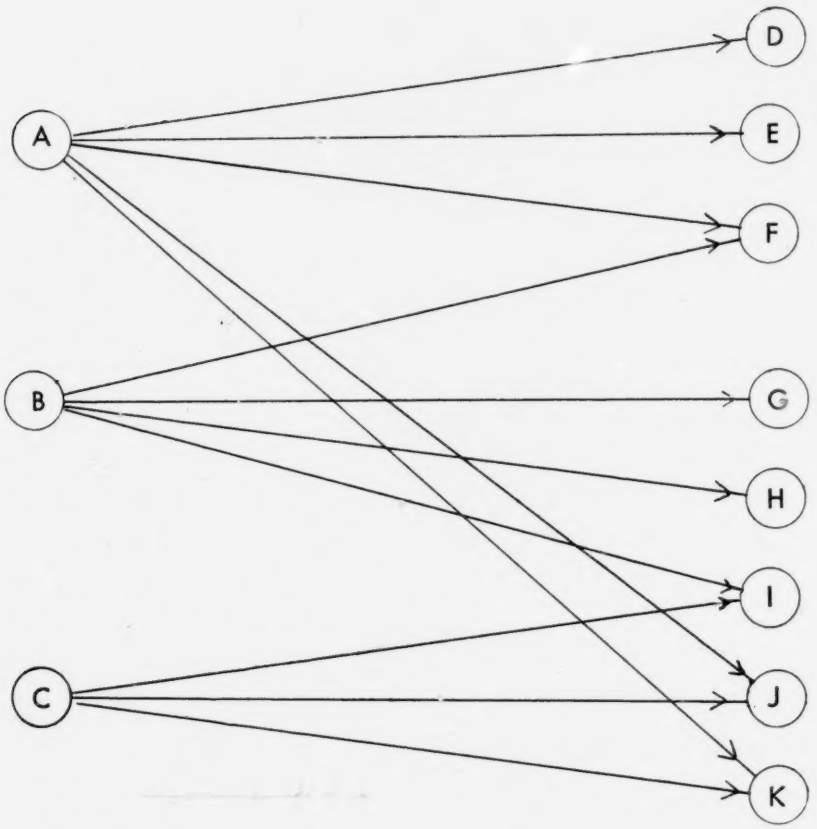
There are various methods for the analysis of citation data which are of potential value for library and information system design. As already mentioned, the simplest method, and that most commonly used, consists of counting citations from given source journals, and constructing a frequency list of cited journals. Such a list may be used as a guide for library holdings. In this type of analysis, the relationship between source and cited journal is ignored once the frequency list has been compiled. The analysis may be described as a one-step process. Figure 1 illustrates in diagrammatic form the stages of the analysis. A, B and C are selected source journals and the cited journals taken from these are represented by D to K. Source journal A cites D, E, F, J and K. Source journal B cites F, G, H and I. Source journal C cites I, J and K. Journals D to K can be ordered into a ranked list according to citation frequency.

An off-shoot of this analysis has been explored by Kessler (1963). He has investigated the relationship of the number of citations in common between articles. He has called the relationship 'bibliographical coupling' and has related the strength of the relationship to the number of citations in common. The source articles related in this manner are assumed to have a similarity of content. In figure 1, A to K can represent articles not journals. Article A cites article J and K which are also cited by article C. Article F is cited by both articles A and B and article I by articles B and C. There is therefore a relationship between all three source articles but the one between A and C is strongest because they have two citations in common - articles J and K. Articles B and C and articles A and B each cite only one article in common.

The third type of analysis may be described as a two-step process. It involves the application of cluster techniques to a frequency list of cited journals. This type of analysis seeks to divide the set of cited journals into subject groupings. The underlying assumptions are that journals which deal with the same subject areas will cite one another. This analysis depends upon the relative frequency of citation of each journal in relation to the source journals. The basic measure by which journal clusters may be judged is the similarity between journals in the same cluster. Once

Figure 1.

ONE-STEP CITATION ANALYSIS



clusters have been identified, journals within a cluster may be arranged in a hierarchy according to citation frequency. Narin(1972) has applied cluster analysis techniques to journals in the disciplines of physics, chemistry and molecular biology, and identified journal groups in subdisciplinary subject areas.

A fourth type of analysis involves the construction of citation networks and may be described as a multistep process. Citations are taken from source journals and used as sources to generate more citations. The relationship between the source and cited journal is significant throughout the analysis. One of the simplest ways of considering a citation network is by a sociometric diagram.[1] In this type of diagram, relationships between a number of references, for example, journal articles, monograph titles, journal titles or authors, are indicated by arrows joining points representing the references as source and cited items;[2] in this way, the distribution of citations and the frequency of citation can be seen clearly.

The present paper, as explained above, is concerned only with groupings of journal articles. A simple citation network is illustrated in figure 2.

The interrelationship between seven reference points A, B, C, D, E, F and G is illustrated. Each point represents a reference consisting of an author and article title. The citation linkages are shown by lines joining the reference points and the direction of the relationship is indicated by an arrow from the source to the cited point. From the diagram, it can be seen that article E is in some way a crucial article because it has more incoming arrows; that is, it receives the greatest number of citations. It represents a great oversimplification of the real situation, as is clearly demonstrated in the trial networks that are reported in Section 3.0.

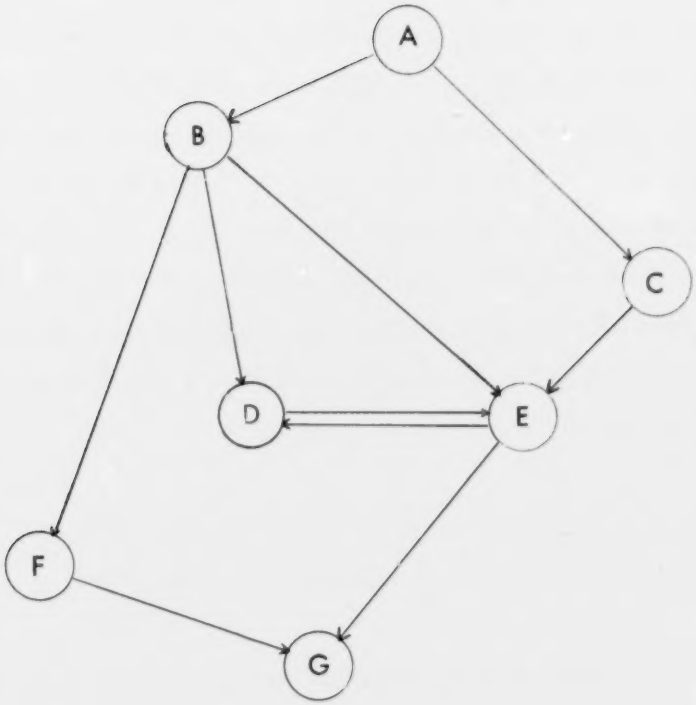
---

1 Sociometric diagram: in the context of bibliographic citations, a diagram indicating the interrelationships between authors or their works.

2 See page 1, footnote.

Figure 2.

A CITATION NETWORK



There are two methods of establishing a network of journal articles. Firstly, it is possible to start with a relatively old article, such as G, and by using a citation index, identify the articles that cite G. This would identify articles E and F. The network can be extended by using the citation index again to find out the articles that cite E and F; this procedure would identify articles B, C and D. These three articles are then checked to see if they have been cited; this produces articles A and articles B and E previously identified. Article A cites both B and C. This method is called a CITED search because it consists of identifying articles that have been cited. The method is described in detail in section 2.2.1.

The second method of constructing a citation network is to start with a relatively recent article and to locate all the relevant citations it contains. These references are then used as sources to generate more citations. This method has been called a SOURCE search and is described in detail in section 2.2.2. In Figure 2 the starting article would be A. Article A makes citations to B and C. Articles B and C are then located and the relevant citations they make are noted. Article B cites D, E and F, article C makes only one relevant citation, to E. Article D makes only one relevant citation, which is also to E. Article E cites article G and article F also cites G. The principle behind this method is that a cited reference is treated as a source reference, and as such it will contain other relevant references, enabling the network to be extended.

The procedure for both SOURCE and CITED searches is repetitive. The iteration is known as 'cycling'; once a source or cited item is located, the procedure is repeated to provide other citation links from the article.

The basic process underlying network construction is the selection of suitable citation linkages that lead to producing the types of network configurations noted above. In making a network of realistic dimensions (in terms of search period, number of references, etc.) it is essential to ensure that the references considered relevant are restricted to those which further the concentration of the network. If every citation was followed up, many 'dead ends' would result. The effort involved in searching all these irrelevant references would be wasted and would not help towards concentrating the citation linkage at particular points in the network. In practice, networks rarely evolve towards such tight concentrations as portrayed in the theoretical cases given,

and where concentration can be developed this is often due to the application of quite rigid cut-offs on the data, e.g. choosing only certain words in an article title, key authors, etc. Determination of the relevance of articles for a network is of great importance and is discussed further in section 2.2.4.

A recently reported study of network construction (Garfield, 1972) on the subject of 'explosive welding' illustrates how citation networks can be used in a search strategy. Although Garfield successfully identifies the key articles in the field he does not state categorically the criteria for choosing citations from source articles which would ensure that the proportion of relevant articles was maximised.

Once a full network has been established, it is relatively easy to see where the concentration of citations takes place and which articles are peripheral, etc., but to begin with it is very difficult to decide suitable cut-off points. For instance, during the construction of a network there may occur many single citations. It is difficult to know whether all of these should be omitted from the final network on the grounds that they do not allow the articles to be grouped, or whether recently published articles should be kept as a separate group, to be further studied, because they may not have had time to be heavily cited. A further problem concerns the relevance of articles in the network. In the case of Figure 2 article C makes a citation to E, but it may contain, say, twelve citations. It is difficult to know beforehand whether journal article C is going to fall in a network that is related to the subject matter at hand. It would be very difficult indeed to establish a network by taking at random any journal article. Some preliminary restriction in the number of cited items, preferably bounded by a subject area, is required. If the search subject is one such as the example given later in this paper ('short-term memory') then journal article C cites E, which is relevant, and may also cite several others not relevant to the network.

In theory the ideal method would be to chart the entire network of citations, with the result that infrequently cited material would not produce concentrations, and could ultimately be discarded. If the data base was extremely large, covering for example all the social sciences or all the physical sciences, then networks would be established in which there were from time to time great concentrations of citations.

The concentrations would represent groupings of articles, where the interrelationship between articles in the groups was strong. These groups may also be weakly related to other groups by infrequently cited articles. A good deal of trial and error is necessary in the first place to ensure that the citations, as they are traced, converge upon a grouping of articles rather than diverge in all directions to produce very weak linkages with an enormous number of articles, perhaps across many subject fields.

2.2 Procedures for retrieval of information using citation data.

There are three methods of constructing citation networks. These have been discussed briefly in section 2.1. Each method is discussed in detail below.

2.2.1 CITED search

This method is adopted when a user wishes to identify articles published subsequently to his chosen starting reference. The method is illustrated by the flow-chart shown in Figure 3.

The method has been called CITED search because it basically consists of locating all articles that have cited the starting reference and then locating articles that have subsequently cited these. A citation index is used for this purpose, Figure 4a shows how o citation network is formed following the procedures outlined in the flow-charts. Each year of the search period is checked to see if the starting reference(s) has been cited. In 1967, article A cited the starting reference. Article A is then checked to see if it has been cited by any relevant articles within the specified search period. It has been cited by articles B and C, and these are considered relevant. Articles B and C are then flagged as LINK references. The term 'LINK' reference has been used because these references provide further citation links if the network is to be extended. Further years of the search period are checked to see if the starting reference has been cited. This may produce article D in 1968. Article D is checked to see if it has been cited subsequently; in fact, it has not. If the citation network is to be extended the articles that have been flagged as LINK references can be recorded and then checked to see if they have been subsequently cited. The network is complete when all the citations which have been flagged as LINK references have been checked to see whether they have been cited in

Figure 3

FLOWCHART OF PROCEDURES FOR THE CONSTRUCTION OF CITATION NETWORKS

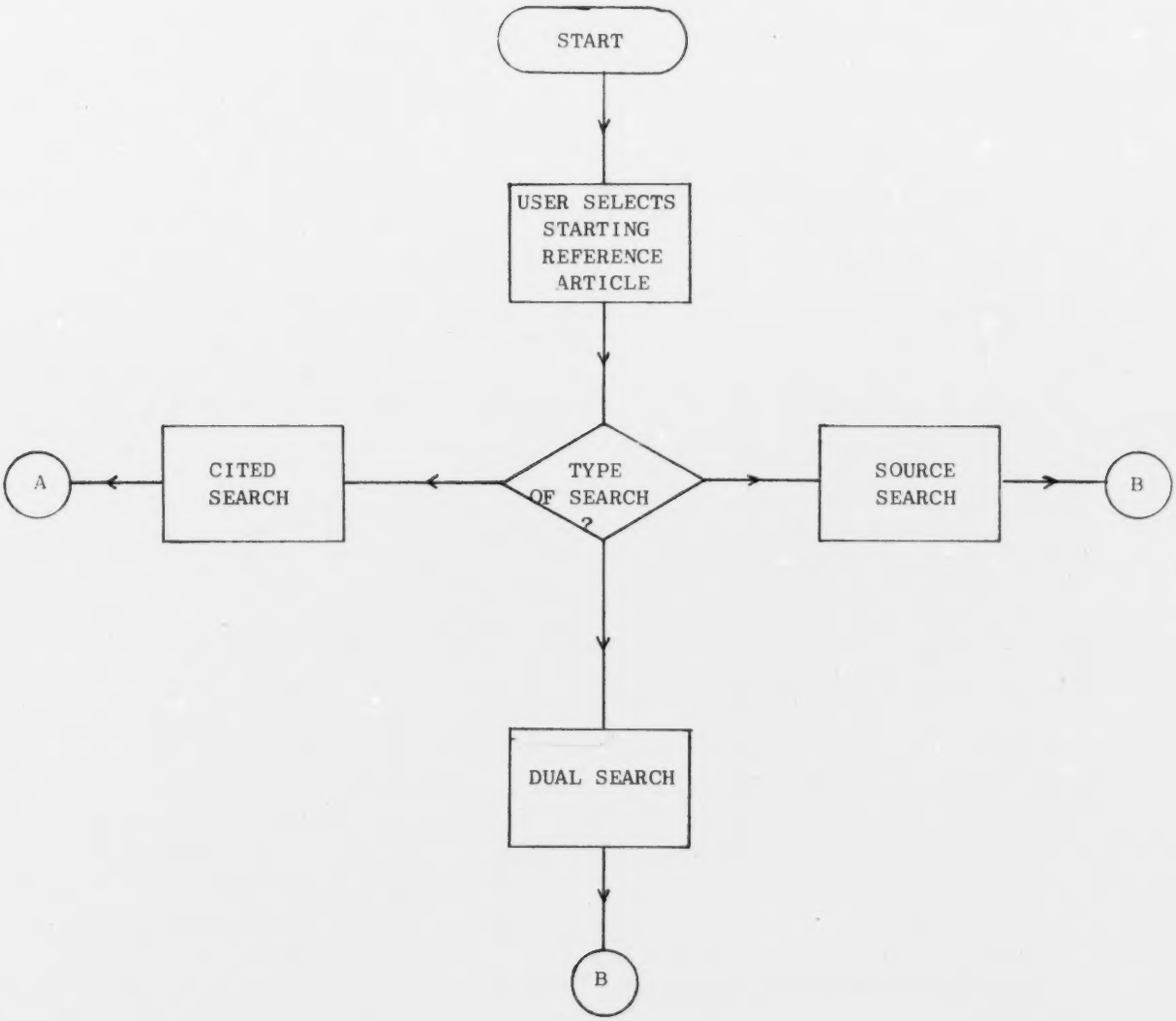


Figure 3 cont'd.

CITED SEARCH

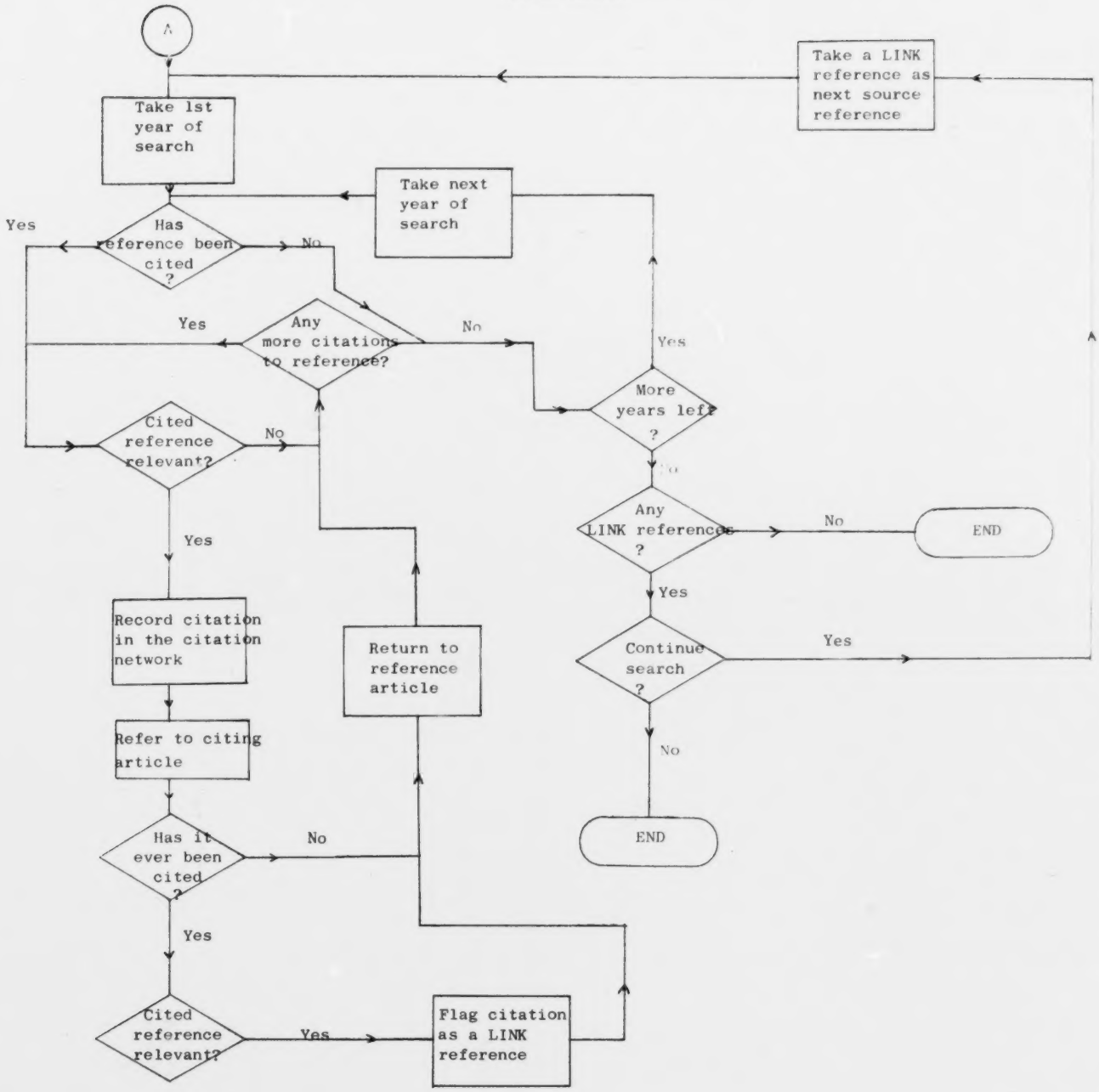


Figure 3 cont'd.

SOURCE SEARCH

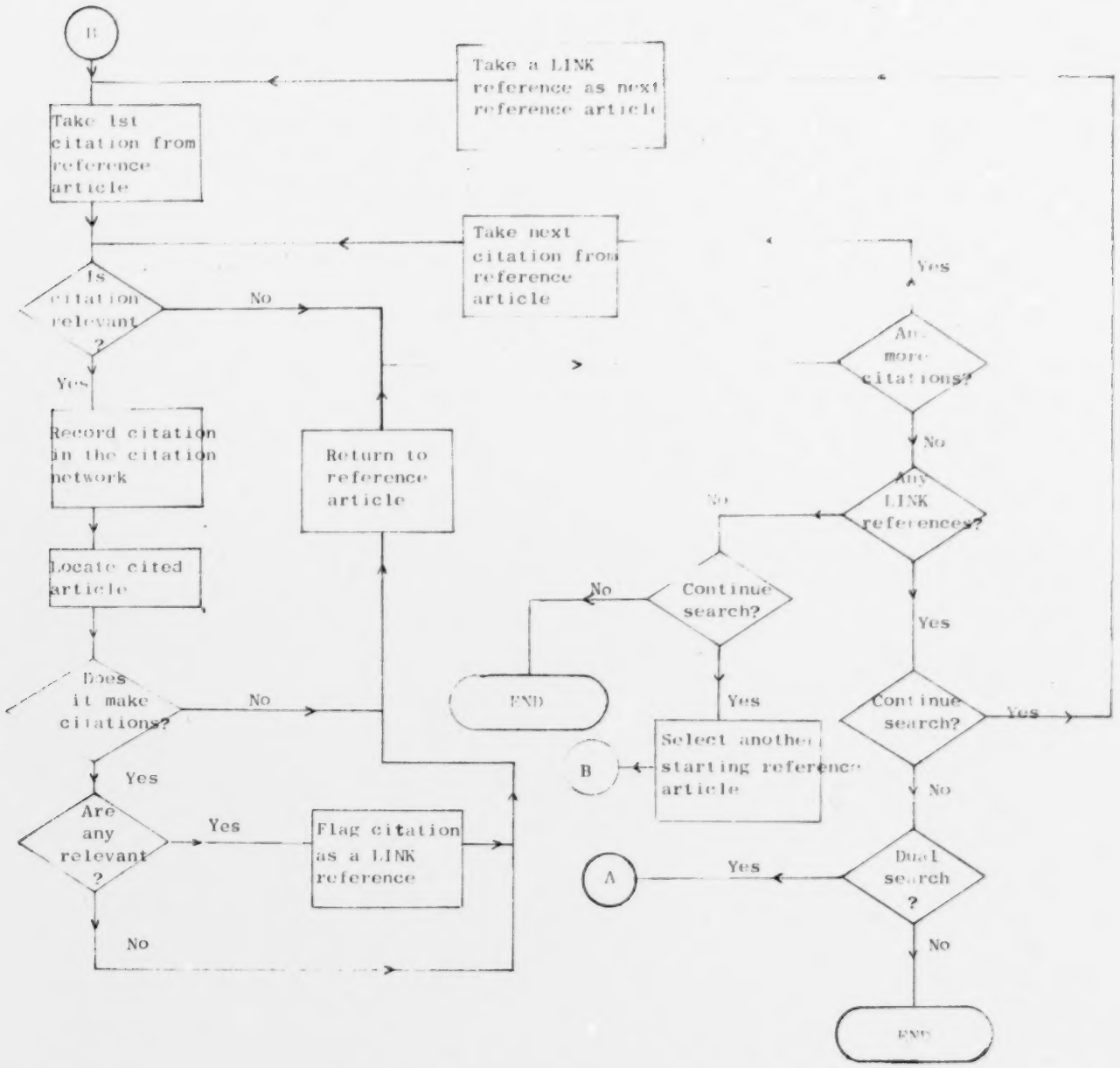
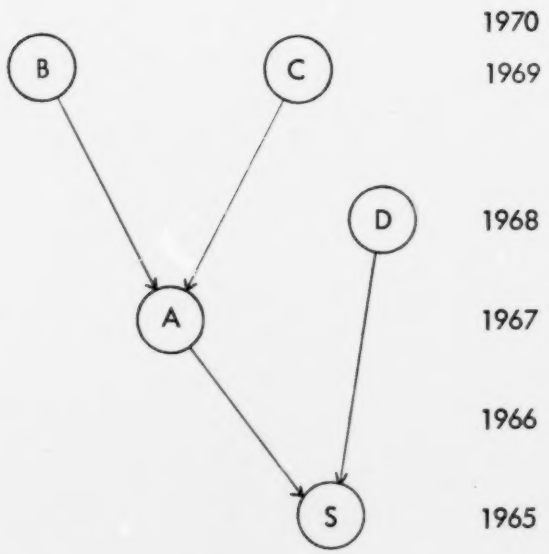


Figure 4a.

NETWORK CONSTRUCTION BY A CITED SEARCH



2.2.2 SOURCE search

This method is adopted when a user wishes to identify a selection of articles published prior to the starting reference. The method is illustrated by the flow-chart shown in Figure 3. The method consists of taking citations from a starting reference, locating them and then using them as sources to provide citations. Figure 4b shows how a citation network is formed following the procedures outlined in the flow-chart. The first citation (A) is taken from the starting reference (S). If the article is relevant it is recorded in the network and the article is located. Citations from A are checked for relevance and the relevant ones are then flagged as LINK references. In the diagram articles B and C are flagged in this manner. The remaining citations from the starting article are then analysed in the same way as A. Article D is considered relevant and recorded, but it does not make any relevant citations. At this point the network cannot be extended, for D may be considered as a 'dead-end'. To extend the network, the citations that have been flagged as LINK references are treated as sources. This process may provide further citations flagged as LINK references. The network is complete when the articles cited by the LINK references are all published outside the dates of the search period.

2.2.3 DUAL search

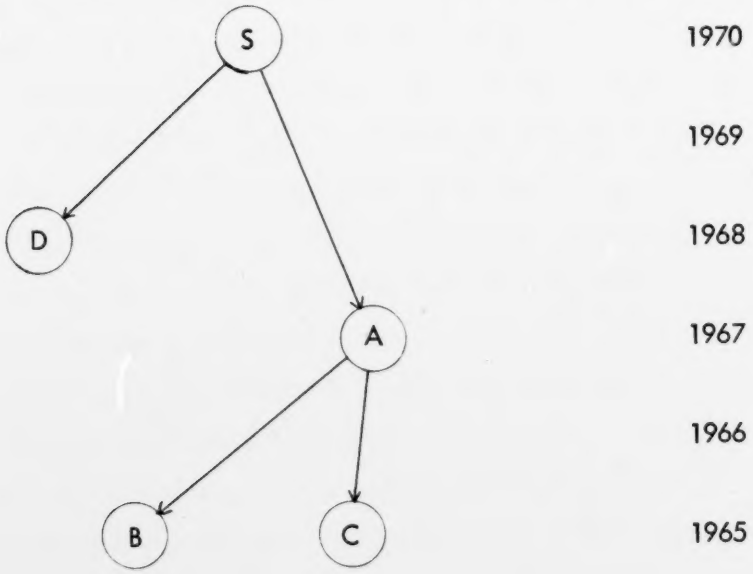
This method uses both SOURCE and CITED methods (hence the name DUAL). The method is adopted when the user wishes to identify a selection of articles which are both prior and subsequent to the starting reference. For example, the user may select a reference published in 1970, but may wish to obtain articles published between 1968-72. If this is the case, a SOURCE search is adopted from 1968-70. All references retrieved in this search are then used as LINK references for a CITED search for 1970-72.

2.2.4. Differences in the procedures for a CITED and SOURCE search

The procedures for conducting CITED and SOURCE searches are different. For a CITED search a citation index can be used. SCI is cumulated into annual volumes, so that each year of the search period has to be taken in turn. However, a five year cumulation has been produced, for 1965-69 and other such cumulations can be expected; these greatly reduce and simplify the search procedure.

Figure 4b.

NETWORK CONSTRUCTION BY A SOURCE SEARCH.



In a SOURCE search a citation index cannot be used. A SOURCE search can only be conducted by locating the actual articles and examining their citations. At present there is no publication that lists citations from given source articles (although of course SCI tapes record this information). This search is not conducted by taking each year of the search period in turn. Instead all citations with the date of publication within the search period are of potential value to the network, and are examined in the order in which they were cited.

2.2.5 Relevance of articles comprising the citation network

Implicit in the method of retrieving information by the construction of a citation network is the assumption that articles contained in it are relevant. The articles are judged relevant by two main criteria: (i) search period, and (ii) subject.

(i) Search period. This is the number of years the user wishes the data base to be searched. During a CITED search the data base is searched for each year of the search period for citations to the starting reference and subsequently for citations to the LINK references. When a SOURCE search takes place the date of each citation that is made from the starting reference and from the subsequent LINK references is checked to see if it is within the search period. If it is, it is considered relevant to the network; if not, it is discarded. The scatter of citations within a network represents their position at that particular time of the network. If the network is extended to cover more years, or reduced to cover fewer years, the scatter of citations and the strength of relationships will change.

(ii) Subject. It is impracticable to trace all citations from all articles, as some articles may cite a large number of items. The network would become extremely large and contain many irrelevant references. Some limiting criteria will have to be used if the network generating procedure is to be practicable. The criteria used should help concentrate the network linkages to a few articles in order to avoid linkages diverging to many articles that are cited once only. As yet there is no clear way of choosing a suitable criterion that works towards this end; an indication of the relevance of articles may be obtained by examining the article title, the author and the journal, but this

procedure is necessarily subjective, and objective criteria would be much more satisfactory.

2.2.6. The starting reference

A starting reference has to be selected. Any reference may be selected so long as it makes citations if a SOURCE search is to be conducted, or is cited if a CITED search is to be carried out. However, the number of irrelevant references in the network is considerably reduced, and the number of relevant references increased, if the reference selected is a key and central work in the field. The number of irrelevant references is also reduced if the starting reference is a review article. However, within limits it does not matter which article is chosen to begin the search, for if enough LINK references are formed, a very similar network may be constructed from any starting reference. The number of irrelevant references in the citation network is also reduced if the date of the starting reference is close to the specified years of the search. This avoids the extension of a citation network outside the search period. For example, if articles are required from 1968-72 and the starting reference is dated 1965, it may be necessary to trace citations through articles in 1965-68 until articles in the search period appear, although articles in 1965-67 will not be required.

Once a network has been constructed the selected starting article does not necessarily assume a special position in the network. The relationship of the starting article to all other articles in the network depends upon the strength of its citation linkages to the other articles. Therefore the fact that one article is chosen to start the search as opposed to another should not affect the final network. This assumes of course that the starting article is relevant to the subject of the search.

2.2.7. Preliminary considerations of the data requirements for construction of citation networks by machine

Any large scale information service based on the production of citation networks requires a machine-readable data base, since several hundred citations may have to be analysed before a meaningful citation network is produced. The problems of compiling such a data base may be enormous. Only a few points are mentioned briefly below.

Firstly, the selection of source journals from which to collect the citations is critical to any information service. This problem is fully dealt with in Working Paper 5. The important point is that any bias in the selection of source journals would be reflected in the citation networks. In general most citations are to a relatively small number of journal titles, articles and authors; the selection of source journals in a broad subject field may produce unforeseen bias. However, in narrow or closely knit subject fields citation patterns may be more various and source journals need selecting with greater care.

All citations from each source journal must be put into the data base. Missing citations may cause gaps in the citation network and cause artificial cut-off points in the network. The number of data fields for each citation must be considered. It may prove feasible only to collect journal title, year, volume, article page numbers and author. Article title may or may not be an important data field, depending upon the type of search the user requires. If a user wishes to search by keywords in title, the article title is obviously required. However, if a straight reference search is required, articles can be uniquely identified by the other five data fields. In fact in some cases it may be desirable for searches to be conducted using the five data fields, for citations do not always give the titles of articles.

A final point is the number of years the data base should span. If retrospective searches are to be carried out the data base should span a fair number of years. Three years should be considered the minimum period for building up a citation network. A period shorter than this is not adequate, as only one link can usually be established between the source and cited item, and no further link can be made from these; it may take a year or more for any article that has been published to be cited.

3.0 APPLICATIONS

Information services based upon citation data have many applications, by virtue of the flexibility which the networking technique offers. The following applications are being considered by DISISS, though not all of them involve actual networks.

3.1 Packages of information

The amount of literature in most fields is growing so quickly that it is very difficult for users to extract and locate the information they require. The problem is eased if information is packaged according to specified user requirements. A package would consist of a list of all the references produced by a citation network.

Such a list could, in theory, be arranged in any manner to meet the requirements of the user. Frequency lists of cited authors and articles could be produced; from these, the user could see at a glance the most important works or authors in the field. He would not have to handle material peripheral to his needs. If required, the package could also contain the citation network; from this the user could determine the relationship of one work with another. The use of citation data provides an ideal method of producing packages of information for groups of users, as the information required may easily be extracted according to specified criteria from any network.

Packages may be developed for particular purposes, not all of them concerned with actual retrieval:

a) To study the occurrence and growth of new concepts.
Packages of information could help to identify new approaches and concepts developing within the social sciences. This may be particularly important for the social sciences, where many theories may exist to explain one particular phenomenon, where schools of thought may die as quickly as they are born, where terminology is unstable, and where there are no 'hard' scientific explanations. The use of traditional indexing services to trace new concepts is particularly difficult in the social sciences because indexing terms can be only assigned when the terms are fully recognised.

b) To study the structure of a discipline and the history of knowledge. A large citation network could describe the structure of a discipline (that is if all the citations within a given subject field were traced for a stated time period). The citation linkages may be considered to indicate the position of an article in the existing body of knowledge. A user may be particularly interested in tracing the evolution of a field; a citation network can indicate the importance of each article in the development of the field. By tracing citations from articles backwards in time (a SOURCE search), the article that first put forward a new concept may be determined. Studies of this kind are more difficult to carry out with a conventional indexing service.

c) To locate information in interdisciplinary areas. The retrieval of information using citation data provides a good method of meeting the information requirements in interdisciplinary fields. Citation networks can be constructed to cut across subject boundaries, provided that the data base contains journals from the disciplines involved.

3.2 Current awareness services

The Institute for Scientific Information (ISI) developed a current awareness system - ASCA (Automatic Subject Citation Alert) - several years ago. It is based on SCI. Each week user profiles are matched with SCI computer entries. If a current article cites any item specified in the user profile, the article is retrieved as being relevant to the user. User profiles are drawn up to include words, phrases, and word stems of article titles, plus author and journal titles.

3.3 Computerised retrieval techniques

The retrieval of information using citation data is suitable for computer based systems. The information systems may be on-line, in which case the user may hold a 'conversation' with the machine in order to retrieve the exact information he requires. As the user has only to provide a starting reference, the problem is avoided of putting user requests in a special form that the machine can use. Equally, batch processing can be used for searches in the same way, though it does not allow 'conversation' with the machine.

3.4 Review series

Citation networks help to indicate the importance of authors and articles in a particular field. This type of data may be of particular relevance for the compilation of review series. Citation networks can delimit the subject field, the key authors and papers; indicate the total field of which the particular review is a subset; provide fairly complete and representative coverage and ensure currency. At the moment there is no way of ensuring that reviews are adequately covering all the relevant material; the review depends very much upon the ability of the reviewer to locate the relevant material and then to make value judgements on the importance of each work and its contribution to the field.

3.5 Foreign language translations

The identification of frequently cited authors and articles may also be particularly important for foreign language translation services. These services may wish to translate only the most important works in a field. Rather than rely on personal expertise, where an element of bias is unavoidable and personal knowledge is bound to be limited, citation data provides some evidence of the works most likely to be wanted in translation.

3.6 Teaching aids

Packages of information may be produced specifically for teaching purposes and as study aids. Groups of articles can be identified that deal with particular aspects of a subject. For instance, the groups could deal with:

a) historical development of a subject
b) a broad outline of the subject
c) detailed aspects of the subject.

Packages may be developed according to detailed specification, so that the information given to a student is directly relevant to his needs. Packages developed in this manner could be used for programmed learning courses.

3.7 Aids to research in information science

The structure of a discipline as revealed by citation networks may provide data on, for example, the shifting of subject boundaries within the social sciences. One traditional means of information retrieval is by classification, but as subject boundaries are not clear-cut or stable, a means of identifying changes is useful.

4.0 PILOT STUDIES

Several pilot studies were carried out to investigate the techniques and procedures for grouping articles using citation data. Science Citation Index (SCI) was used for the studies; because it does not contain sources published before 1964, it was necessary in some studies to use source journals not included in SCI in order to complete the study.

4.1 Studies to identify groups of articles with keywords in the title

The main objective of these studies was to retrieve articles with chosen keywords in the titles and to identify the core articles and the peripheral articles on the topic. The keywords chosen were 'short-term memory' and 'deviant behaviour' or synonyms. A citation network was constructed for each using citation frequency data.

4.1.1 Short-term memory - a DUAL search

The objective of this study was to retrieve articles between 1959 and 1970 with the keywords 'short-term memory' or synonyms in the title of the article. The resulting network is shown in Figure 5. The starting reference was no. 1. The 1969 SCI[1] was used to identify any relevant citations made in 1969 to reference no. 1. The procedure was that of a CITED search. Twenty-two citations were identified as relevant and these are indicated by nos. 36-57 in Figure 5. The network was extended by using a SOURCE search to identify any relevant articles cited by the twenty-two articles. This procedure located the articles numbered 2-35. Citations from these articles were not checked because the number of references compiling the network was becoming too unwieldy to deal with by hand. The resulting network is very incomplete, being based only on citations made in 1969 to the starting reference. Ideally, each of the years spanned by the network should be checked for citations to the starting reference and the citations arising from these should similarly be checked for relevance.

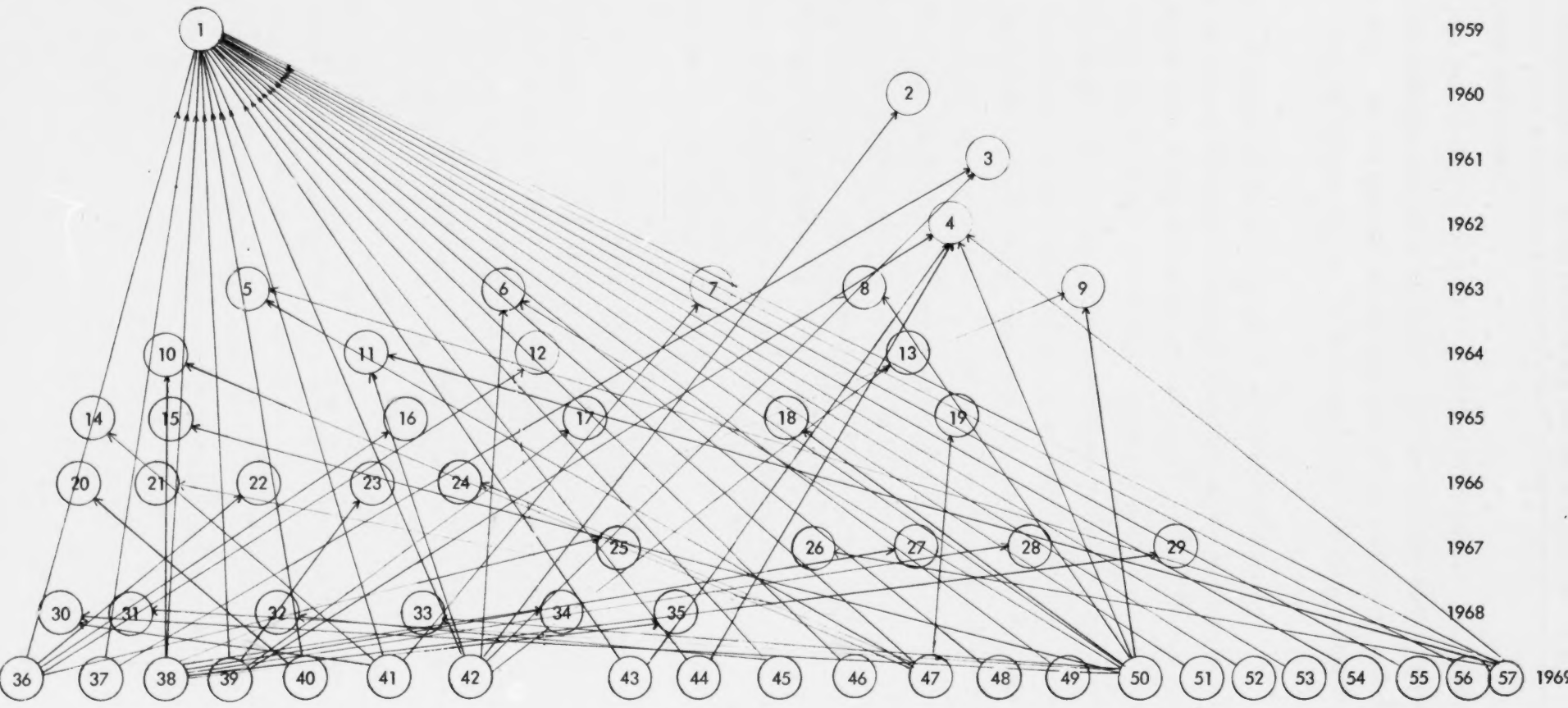
The resulting citation network was presented for evaluation to a person who knew the subject well. He felt that the articles indicated by the citation network as being important were important articles in the field. The user was also able to divide the references into distinct subject groupings. These are given below. The subject groups are not mutually exclusive.

---

1. The 1969 edition was chosen because this was the earliest year of SCI in Bath

Figure 5.

SHORT-TERM MEMORY CITATION NETWORK



References to this network are given in Appendix A.

(i) paired associate learning. References 3, 6, 7, 8, 21.

(ii) interpolated activity/inertial interval. References 15, 16, 18, 25-30, 36, 38-41, 48-50, 53-55.

(iii) semantic acoustic confusability. References 10, 19, 24, 43, 52, 56.

(iv) proactive/retroactive interference. References 1, 2, 4, 5, 9, 11, 12, 15, 17, 19-23, 31, 33, 44, 45, 47, 55-57.

The user was not familiar with five of the references, 13, 14, 32, 42 and 51, so he could make no value judgements on these. Four references, 34, 35, 37 and 43, were thought to be of marginal relevance to any of the four subject groups identified.

Although the articles in the network could be classified into subject groups, the citation linkages in the network were not necessarily between articles within the same subject group. The starting article concerned proactive/retroactive interference and it was to be expected that the number of articles contained in this subject group would be greater than in the other groups. To illustrate the exact nature of the citation linkages the group on proactive/retroactive interference will be examined in detail. The starting article, reference 1, was cited by twenty-two articles in 1969, but only five (nos. 44, 45, 47, 56, 57) were considered as being concerned with proactive/retroactive interference. Three of the five articles together made seven citations to the articles besides reference 1 in the subject group listed below.

From reference 44 : to reference 34.
From reference 47 : to references 5, 19.
From reference 57 : to references 4, 5, 11, 26.

Within the subject group four of the citation links were to two articles only; references 4 and 5 were both cited twice. This is an indication of concentration of citation links on specific references. If the network had been extended by tracing the citation links through all years (1959 - 1969) it is envisaged that the citation links within subject groups would be more evident and further concentrations of citation links would occur, as each article would have a greater chance of being cited.

The subject group with the next largest number of articles in it is that on interpolated activity/inertial interval. Citation linkages do occur within this group; they are listed below.

From reference 36 : to references 16 and 25.
From reference 38 : to references 27, 28, 29.
From reference 41 : to reference 30.
From reference 50 : to references 15, 18, 30.

Only one article is cited twice from within the group. This is reference 30, cited by references 41 and 50.

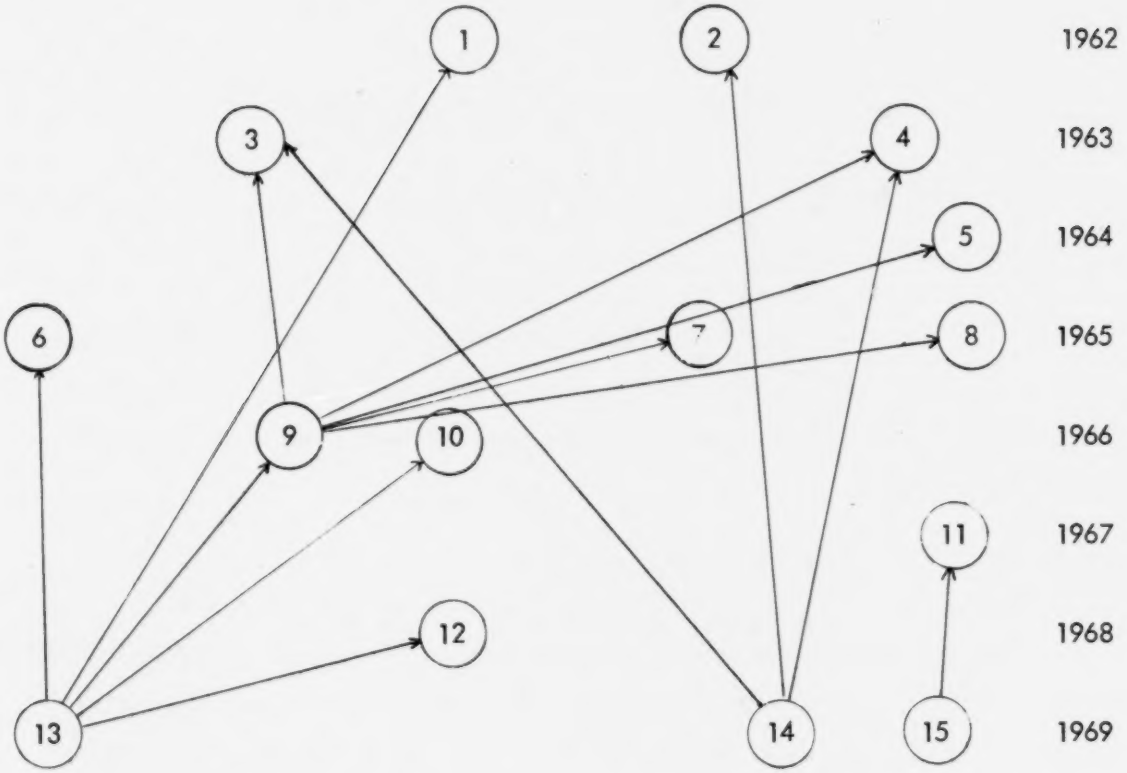
There are only five and six articles respectively in the remaining two subject groups - paired associate learning and semantic acoustic confusability. No citation linkages occur within either groups and this can be attributed to the small number of articles in the groups.

4.1.2. Deviant behaviour - a SOURCE search

The main objective of the study was to retrieve articles between 1962 and 1969 with the keywords 'deviant behaviour' or synonyms in the title of the article. The starting reference was 13 in Figure 6. A SOURCE search was first carried out. The starting reference made five citations to other works that were relevant for the search. These are references 1, 6, 9, 10 and 12. Each of these references was checked for any relevant citations it might make. The citations made in reference 1 all went outside the years of the search period. Reference 9 made five relevant citations. These are references 3, 4, 5 7 and 8. Each of these references was then checked for citations, but none contained any relevant citations within the search period. To increase the number of articles in the network the term 'deviant behaviour' was looked up in the Permuterm Subject Index for 1969; two articles were identified, references 14 and 15. Reference 14 cited reference 2 and reinforced citation linkages to references 3 and 4. Reference 15 made only one relevant citation, to reference 11.

Figure 6.

'DEVIANT BEHAVIOUR' CITATION NETWORK



References to this network are given in Appendix B.

The total network was composed of 15 references only. The small number may be due to the fact that the term 'deviant behaviour' is very broad and article titles may use more specific terms, or they may use other terms for the same subject. The search would no doubt have located more references if all the citations from each article had been used, instead of only the citations that had the keywords in the title.

The network was presented for evaluation to an academic researcher working in the field of deviant behaviour. He judged that most of the articles that occurred in the network were important in the field. Reference 14 was not known to the user, although he thought it would be relevant. The reason he had not come up against the reference before was that it was in the Journal of Mental Hygiene, a journal that sociologists do not usually scan. The starting article gave a general treatment to deviancy and the majority of references retrieved also treated the subject in this manner. Only references 8 and 18 dealt with more specific aspects of deviancy.

4.2 Studies to identify articles containing new concepts, new ideas and new terms.

The main objective of these studies was to see how well an information retrieval procedure based on citation data would perform in handling literature that dealt with new concepts, new ideas or new terminology in the social sciences.

Fifteen topics were chosen at random as terms that have been introduced into social science literature over the past ten years. Each term was looked up in ISI's Permuterm Subject Index for the years 1969, 1970 and 1971, and the number of articles with the term in the article title was recorded. The results are shown in the following list.

	1969	1970	1971*
Autistic children	16	12	3
Autokinetic	7	16	5
Biofeedback	0	0	1
Bionics	6	1	2
Cognitive controls(s)	3	4	2
Generation gap(s)	8	5	12
Orienting response	13	12	8
Phillips curve(s)	0	2	0
Preparatory response(s)	1	3	0
Representational processes	0	0	0
Skin resistance(s)	21	12	9
Synectics	2	1	2
T-groups	29	1	4
Transcendental meditation	0	2	0
Yoga	1	3	0

* Data available only for January to September 1971.

The list indicates in a very preliminary way the frequency of use of the terms in the literature. Terms that occur infrequently in all 3 years can be interpreted as being of fairly recent origin as showing that only a few people were publishing work in that particular field. Terms occurring frequently in 1969 and declining in 1970 and 1971 may be interpreted as being of declining importance in the literature. An example of this appears to be T-groups. However as the data was only available over three years (only these three volumes of the Permuterm Subject Index were available in Bath University Library) firm conclusions cannot be reached at this stage. If data were available over a period of ten years the conclusions would have more validity. A further limitation of the present study is that it is based only on ISI's files; subject coverage of some of the terms in the above list may not be comprehensive.

The procedure for identifying the articles in the studies was as follows. The articles occurring in the 1971 Permuterm Subject Index were located in the Source Index to obtain full bibliographical details and to obtain the number of citations each reference

made. It was decided to work from the most recent data available, namely 1971. Those that did not make any citations were discarded from the study for they could not provide links to other articles. The references with the greatest number of citations were located first. Ideally a total citation network should be constructed by tracing all citations from every article. This is a large task for a manual operation; the procedure adopted was to identify first the citations to the oldest material, the assumption being that these articles were the ones most likely to lead to the articles that first put forward the new concepts. The basic procedure used was that of a SOURCE search described in section 2.2.2. It was modified by selecting the earliest citations from the articles instead of tracing citations from all articles. To validate the choice of citations used in the network the text of each article was read, and this in some cases indicated the earliest relevant citation. As the terms were traced back an indefinite number of years, the studies invariably went outside the scope of ISI's data base. Each article was obtained and the citations from it were scanned. If the publication was not available in Bath University Library, a request was made to the National Lending Library for Science and Technology. Sometimes a publication was unobtainable because it was very old and insufficient bibliographical details were available to enable it to be located. When this occurred the network was artificially cut off and work could not proceed further unless other publications in the network led to the publication that could not be located.

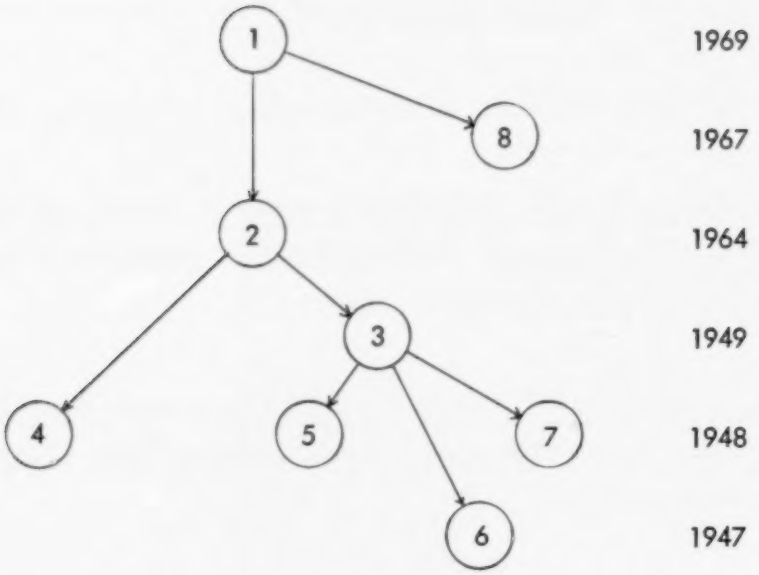
The results of tracing five terms are given below:

a) T-groups

Reference 1 in Figure 7 was chosen as the starting point in the citation network because it was a review article and made 75 citations. Analysis of the citations by title revealed the ones that would be most likely to give leads to earlier works on the T-group concept. When these publications were obtained it was found that only two were directly relevant. These references 2 and 8 both describe the historical development of the concept. References 5 and 6 are the earliest publications dealing specifically with 'training groups'. Reference 7 indicates the historical development of the concept. Reference 3 is an important publication in the field for it describes the first experiments

Figure 7.

T-GROUPS CITATION NETWORK



REFERENCES TO FIGURE 7.

1. GOTTSCHALK, L. A. & PATTISON, E.M. Psychiatric perspectives on T-groups and the laboratory movement: an overview. American Journal of Psychiatry, 1969, 126, 823-839.

2. BRADFORD, L. P., GIBB, J. R. & BENNE, K. D. T-group theory and laboratory method. New York, Wiley, 1964.

3. LIPPITT, K. Training in community relations. New York, Harper, 1947.

4. BENNE, K. D. Principles of training method. The Group, 1948, 10 (2), 17.

5. Proceedings of the Second Summer Laboratory of the National Training Laboratory in Group Development. Washington D.C., Department of Adult Education, National Education Association, 1948.

6. Proceedings of the First National Training Laboratory in Group Development. Washington D.C., Department of Adult Education, National Education Association, 1947.

7. BRADFORD, L. (ed.). A tale of three conferences. Adult Education Bulletin, 1948, 12(3).

8. BRADFORD, L. P. Bibliography of an institution. Journal of Applied Behavioural Science, 1967, 3, 127-143.

with T-groups and hence cited the reports from the laboratory. The authors most frequently cited appear to be L. P. Bradford, K. D. Benne, and R. Lippitt; these men could be considered the innovators in the T-group field.

b) Short-term memory

Further work was done on the short-term memory citation network (Figure 5) to see if the first appearance of the term could be detected. Using the SCI no other article earlier than Peterson & Peterson, reference 1, could be found with the keywords in the title.

c) Phillips curve

Only two references were given in the Permuterm Subject Index. Neither of these made any citations, and so the study could proceed no further for this term.

d) Synectics

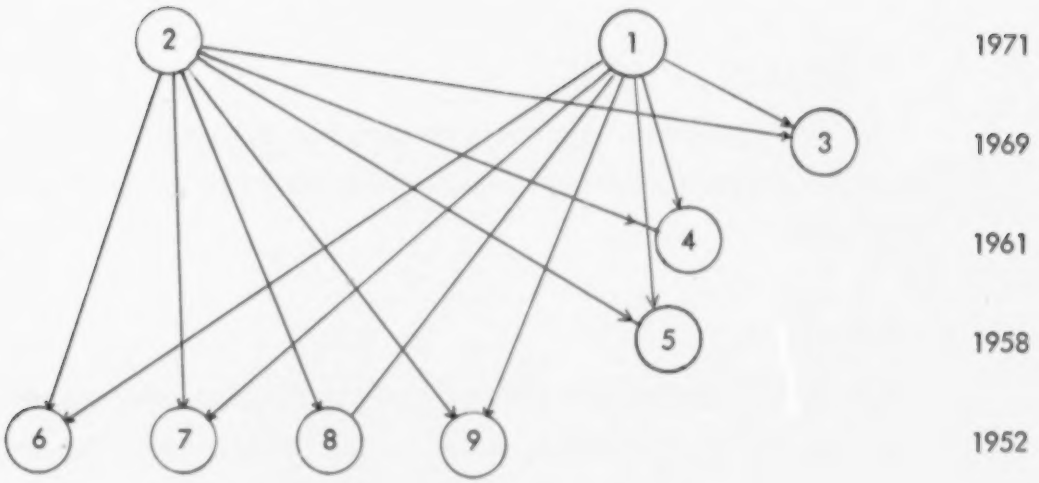
The two references made in 1972 were both by the same author and both articles made almost the same citations. The citations used in the network were common to both articles. References 1 and 2 cited several papers by S.J.J. Gordon (6) published by Arthur D. Little, Inc., but no dates of publication were given. Arthur D. Little came into operation in 1952 so the papers must post-date this. From the text of reference 4 it may be assumed that those papers were published in 1952 and deal with the results of the research started in 1944. The first appearance of the term 'synectics' is in the publication of S.J.J. Gordon, 1961, reference 4.

e) Generation gap

Only a few references located in the Permuterm Subject Index made citations. Eight, five and twelve references were located respectively in the 1969, 1970 and 1971 Permuterm Subject Index, but only references 1, 2 and 4 made any citations. The reference making the most citations, and also citations to the oldest material, was Murray 1971. This article was chosen as the starting point; of its 33 citations three were identified as being closely linked. These are shown in Figure 9.

Figure 8.

SYNECTICS CITATION NETWORK



REFERENCES TO FIGURE 8.

1. RAUDSEPP, E. Synectics - enhancing group creativity, 1 Hydrocarbon Processing, 1971, 49, 105.

2. RAUDSEPP, E. Synectics-enhancing group creativity, 2 Hydrocarbon Processing, 1971, 50, 133.

3. PRICE, G. M. The practice of creativity. Cambridge, Massachussetts, Synectics Inc., 1969.

4. GORDON, S. J. J. Synectics. New York, Harper & Row, 1961.

5. WILSON, W. Operational creativity. Michigan, Bureau of Business and Economic Research, College of Business and Public Service, Michigan State University, 1958. (Marketing and transportation paper no. 2.)

6. GORDON, W. J. J. A preliminary examination of operational creativity. Arthur D. Little, p. 10. (Mimeoed).

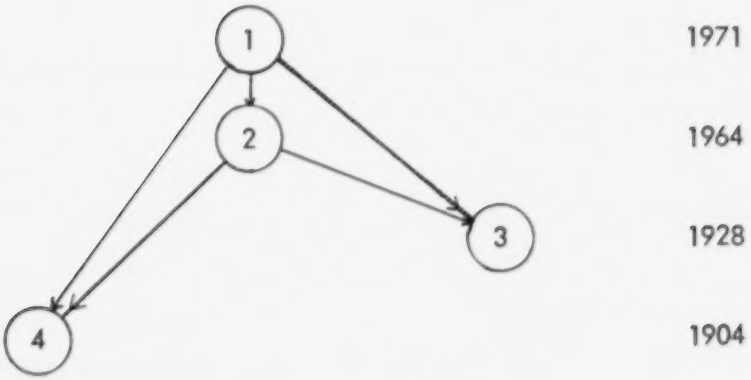
7. GORDON, W. J. J. Invention in art and engineering. Arthur D. Little, p.5 (Mimeoed).

8. GORDON, W. J. J. The integration of creative persons. Arthur D. Little, p.11. (Mimeoed)

9. GORDON, W. J. J. Some environmental aspects of creativity. Arthur D. Little, p. 6. (Mimeoed).

Figure 9.

GENERATION GAP CITATION NETWORK



REFERENCES TO FIGURE 9.

1. MURRAY, J. B. The generation gap. Journal of Genetic Psychology. 1971, 118, 71-80.

2. GRINDER, R. E. & STRICKLAND, C. E. G. Stanley Hall and the social significance of adolescence. Teacher's College Record, 1964, 64, 390-399.

3. MEAD, M. Coming of age in Samoa. London, Cape, 1928.

4. HALL, G. S. Adolescence. New York, Appleton, 1904.

References 3 and 4 are both cited by references 1 and 2. From the citation pattern, it appears that the term 'Generation gap' is associated with early studies of adolescence.

This study was investigated further in order to establish how closely the term 'Generation gap' is associated with early studies of adolescence.

Two approaches were taken:-

i) tracing citations to Mead, M. 1928 in the 1969, 1970 and 1971 SCI, and

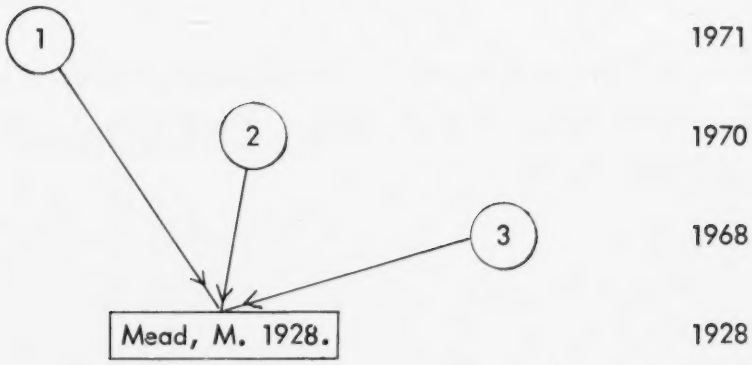
ii) tracing the citations from the two articles identified in the 1971 Permuterm Subject Index.

The resulting networks are shown in Figures 9a and 9b.

From Figure 9a, it can be seen that only 2 authors cited Mead 1928; Mead, F. made a citation in 1968 but it was entered in the 1969 SCI. The third author, J. Ablon (1970), cited the revised 1961 edition of the 1928 work. The network shown in Figure 9b was constructed using the SOURCE search method described in section 2.2.2. The 33 references cited by Murray (reference 1) and the four references cited by Thomas (reference 2) were obtained. The citations that these articles made were checked against each other; only the ones in common are shown in Figure 9b. It was too large a task to construct by hand a network containing all the citations from reference 1, as some of these articles contained over 70 citations. A good proportion but not all of the articles cited by Murray could be obtained by the time this paper was written. An interesting feature of the citation pattern is that in 1967 articles cited other articles published in the same year. These are references 7, 9, 10 and 11. All these articles occur in fact in the same issue of the same journal. The authors must have been in close contact with each other's work.

Figure 9a.

1969, 1970, 1971 CITATIONS TO MEAD, M. 1928.



REFERENCES TO FIGURE 9a.

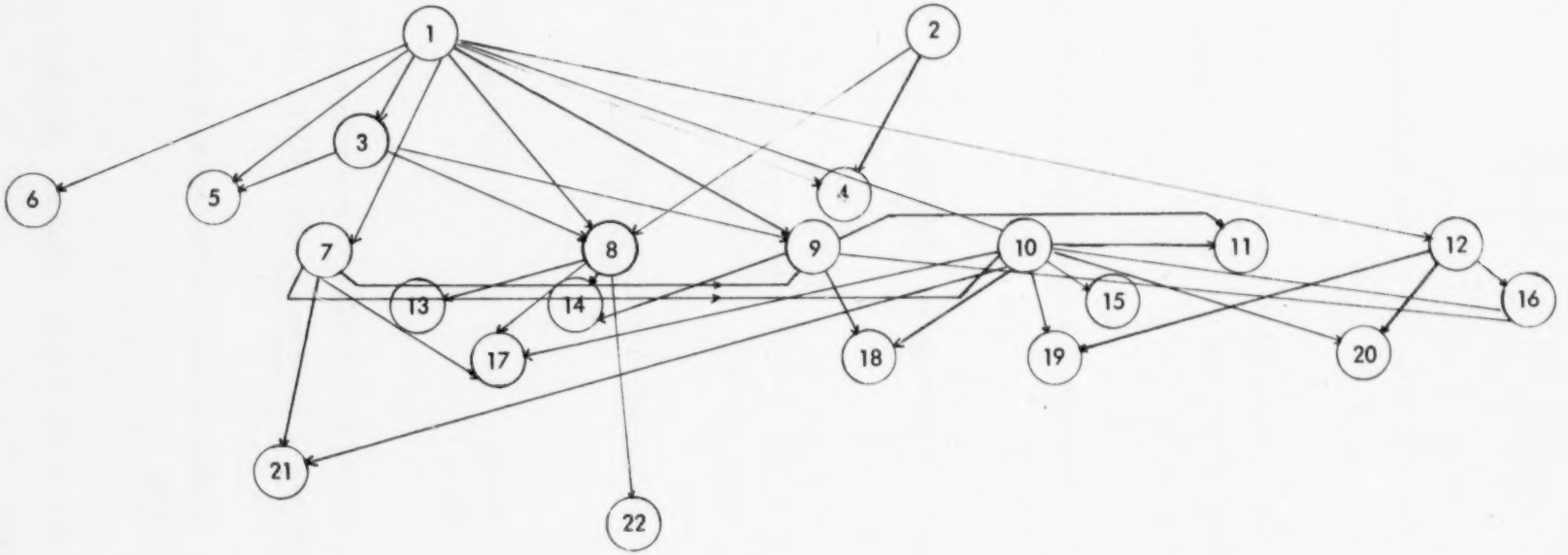
1. MUNROE, R. H. &MUNROE, R. L. Household density and infant care in an East African Society. Journal of Social Psychology, 1971, 38, 3-14.

2. ABLON, J. Samoan funeral in urban America. Ethnology, 1970, 9, 209.

3. MEAD, F. Observations on ecology and behaviour in a natural environment of Tylos Latreillei audoum (isopoda tylidae), Vie et Milieu, Serie C, Biologie Terrestre, 1968, 19, 345.

Figure 9b.

GENERATION GAP CITATION NETWORK EXTENDED



REFERENCES TO FIGURE 9b.

1. MURRAY, J. B. The generation gap.. Journal of Genetic Psychology, 1971, 118, 71-81.

2. THOMAS, L. E. Political generation gap; a study of liberal and conservative activist and nonactivist students and their parents. Journal of Social Psychology, 1971, 84, 313-314.

3. RUBINSTEIN, E. A. Paradoxes of student protests. American Psychologist, 1969, 24(2), 133-141.

4. FLEUR, L. S. Conflict of generations. Saturday Review, 1969, January 18, 53-55, 66-68.

5. ABRAHAMS, M. H. et al. Student protests: a phenomenon for behavioral sciences research. Science, 1968, 161 (3836), 20-23.

6. KENISTON, K. Youth change and violence. American Scholar, 1968, 38, 227-245.

7. JENKS, C., AND REISMAN, D. The war between generations. The Record, 1967, 69(1) 1-21.

8. FLACKS, R. The liberated generation; an exploration of the roots of student protest. Journal of Social Issues, 1967, 23(3), 52-75.

9. SAMPSON, E. E. Student activism and the decade of protest. Journal of Social Issues, 1967,23(3), 1-33.

10. KENISTON, K. The sources of student dissent. Journal of Social Issues, 1967, 23(3), 108-137.

(References to 9b continued).

11. BLOCK, J. H. et al. Activism and apathy in contemporary adolescents. In: J. F. Adams (ed.), Contributions to the understanding of adolescence, New York : Allyn & Bacon, 1967.

12. TRENT, J. W. &CRAISE J. L. Commitment and conformity in the American college. Journal of Social Issues, 1967, 23(3), 34-51.

13. LIPSET, SEYMOUR & ALTBACH, P. Student politics and higher education in the United States. Comparative Education Review, 1966, 10, 320-349.

14. LIPSET, S. University students and politics in underdeveloped countries. Comparative Education Review, 1966, 10, 132-162.

15. KENISTON, K. The faces in the lecture room. In: R. S. Morison (ed.), The American university. Boston : Houghton Mifflin, 1966.

16. PETERSON, R. E. The scope of organised student protest in 1964-65. Princeton : Educational Testing Service, 1966.

17. KENISTON, K. The uncommitted. New York : Harcourt Brace, 1965.

18. MICHAEL, D. The next generation; the prospects ahead for the youth of today and tomorrow. New York : Vintage, 1965.

19. DRAPER, H. Berkeley, the new student revolt. New York: Grove, 1965.

20. SOMERS, R. H. The mainsprings of the rebellion : a survey of Berkeley students in November, 1964. In: S. Lipset & S. Wolin (eds.) The Berkeley student revolt. Garden City, New York : Doubleday 1965.

21. MILLER, D. & SWANSON, G. E. The changing American parent. New York : Wiley, 1958.

(References to 9b continued).

22. EISENSTADT, S. N. From generation to generation. Glencoe: The Free Press, 1956.

It was hoped that a comparison of the networks shown in Figures 9a and 9b might show some articles in common, but this was not so. The articles in the network shown in Figure 9b deal mainly with 'student protests'.

4.3 Pilot Studies: Problems and Conclusions

The use of citation data to identify groups of articles proved to be successful. The study on short-term memory (section 4.1.1.) showed four subject groupings of articles with citation linkages present within two of the groups. There was evidence of the concentrating of the citation linkages at particular references within the groups. If the study had been extended and all citations had been traced through all years of the network rather than 1969, citation linkages within all the groups would have increased.

The study using the keyword 'deviant behaviour' (section 4.1.2) did not show such clear groupings of articles. If the network had been constructed without the restriction of the keyword, groups of articles would no doubt have appeared. The use of keywords in this study was too limited, the keyword chosen was too broad for groups of articles to be identified by title. However, these studies were purposely conducted using keywords in titles, because it was felt that they would reduce the number of references retrieved and would therefore be more suitable for manual handling. If all the citations from each article had been used in the study, the network would have become too large.

The studies to identify articles containing new concepts (section 4.2) showed that although early articles may be identified, there is no firm evidence to suggest exactly which article first put forward the new concept. This is due to the fact that when constructing the networks it was difficult to define the relevant articles; in the study it was only the earliest citation made by each article which was considered relevant.

Previous studies on tracing the history of subjects by citation linkages have not put forward any objective measures for establishing the relevance of articles. Garfield has reported several studies on the use of SCI to trace the history of various subjects.

However his studies were primarily aimed at identifying the important articles in the development of the subject rather than identifying articles that first put forward a new concept. It may be that some articles that first put forward a new concept are considered important articles in that field but it may not be necessarily true for all of them. This may account for the fact that no firm conclusions could be reached in the pilot studies. It may be illuminating at this point to examine the methods which Garfield used to trace the history of particular subjects.

In 1964 Garfield, Sher and Torpie identified the key DNA discoveries as described in Isaac Asimov's The Genetic Code, and then carried out a literature search using conventional bibliographic tools in order to identify the articles that corresponded to the historical events as described by Asimov. The citation linkages between each article were examined to see how far the linkages represented the historical events. The study concluded that citation patterns were a valid means of investigating historical events. In 1969 Garfield reported a study on the recent history of DNA. From a given list of 30 to 40 articles published in 1967 on the subject of DNA he compiled a master list of all citations in their articles. He disregarded all articles in this list cited less than five times. He then checked the list of articles in SCI to make sure that they were all highly cited articles. He then repeated the process for all years of the study 1961 to 1967. In this manner he claimed to have located all the most important articles on the recent history of DNA. However, Garfield does not explain how he chose the thirty or forty articles for each year of the study. The choice of his starting articles is critical to his method, and although all unimportant articles are eliminated by checking in the SCI, some important articles may not have been cited by the starting articles. The method does not provide a citation network as only one citation link is used to provide further articles in each year of the search. The method is that described as a one step citation analysis in section 2.1. The choice of the 30 or 40 starting articles is therefore critical in this case because any bias in the selection of the starting articles cannot be balanced out by cycling through all years of the search period. A cycling procedure would provide further citation linkages which would reaffirm the importance of an article previously identified.

More recently, Garfield (1970) conducted a study on the history of the design of electromagnetic flowmeters in order to identify 'edifying' articles. The method used was to construct a citation network. A known article was selected and the citations it made examined. The relevance of each article was determined by examining authors, titles, citations, and frequency of citation as indicated in SCI.

The Permuterm Subject Index was also checked for relevant articles and altogether about 500 articles were produced of which 159 were relevant. The final network contained twenty-six articles. These were identified by using SCI; the greater the number of citations to an earlier work, the greater was the likelihood that the cited paper was a key event in the subject field. In this study Garfield used a citation network to identify the key articles but he gives little detail of the precise procedure he followed and of the amount of cycling he had to do.

The outstanding problem revealed by the pilot studies is the successful identification of relevant citations that enable the networks to be constructed in the shortest possible time; ideally the articles identified must lead to a concentration of the network, rather than to a scatter of references that have only been cited once. Garfield (1970), who has done perhaps most work on use of citation networks, does not appear to put forward any definitive method of selecting the appropriate articles. He suggests the following criteria: author, title, citations made by the article, and finally the number of times the article has been cited in SCI.

These criteria are not very satisfactory for the following reasons. An author can be used as a criterion only if he is known to the user as having worked in the field. Key authors may be easy to recognise; but problems arise with the large proportion of lesser known authors. The title of an article may certainly be used as an indication of relevance but serious misjudgements may occur. The citations an article makes may give some indication of relevance, but the article must first be located for the citations to be scanned, and this may be a time consuming task. Finally, Garfield suggests checking in SCI - this would give an indication of the importance of the work. But if all or some of these procedures have to be carried out for every citation made by the starting article (some articles may make up to 60 citations or even more) and then all the preceding articles, the retrieval process must be extremely slow.

Further problems in constructing networks concern the size and structure of the data base from which the network is compiled. For an effective information service, the citation data base needs to be very large if a subject or discipline is to be well covered by the service. Not only that, but it must be composed of a valid selection of source journals. If the service is to be economically viable it may be necessary to make a selection of journals. An integral part of DISISS research is to investigate the application of cluster analysis techniques to journal selection. Clustering techniques

are being used to identify groups of frequently cited journals. Journals selected in this manner would provide the basis of an information store for a service. Retrieval of information from such a service could be carried out by using the methods described in this paper, based on citation data.

Other problems that are of relevance to citation networks concern the nature of citation practice. Several papers have been concerned with these problems; see Garfield and Sher (1963) East and Weyman (1969), Price (1965). The main areas for concern are mentioned briefly below.

The major assumption about citation practice is that the contents of the article will determine which references an article will cite. However, an author may cite a work for other reasons. Some material may be cited because it is so widely established that to ignore it would be an omission, whether it was used or not. An author may cite material that he has not read, perhaps in order to give more weight to his own work. An item may be cited because it supports a particular point of marginal relevance to the main theme of the paper. An author may wish to make his own previous work more widely known, and he may therefore cite it without its being of strict relevance to the present work. Finally, there is no way to ensure that the author was aware of all the relevant articles when he wrote up his work, nor that he cited all works that went towards the writing of his own article. Although citations are by no means a perfect indicator of use, deficiencies diminish in importance as a network grows in size, and individual inadequacies tend to cancel out. In a field such as social work, where citations are less frequent, and where less effort may have been made by an author to search previous literature than in a 'pure' research field, the problems may be much greater.

5.0 FUTURE WORK

DISISS will continue to investigate the use of citation data for the identification of groups of articles. It is hoped that it may prove feasible to develop information services based on citation data. In particular DISISS will look at the use of citation data to construct packages of information, also its use in the selection of foreign language items for translation. The problems that have been discussed in this paper will be further investigated. Special attention will be directed towards the establishment of suitable criteria for the determination of the relevance of articles for a citation network. It will not be possible to generate effective citation networks from present DISISS citation data files due to the fact that data for citation networks must be derived from all source journals for each year the network is to span.

At present the DISISS citation data files are made up as follows. Data collected in the citation pilot study is composed of citations from source journals published in 1950, 1960 and 1970. All citations were collected from every third article. The data collection for all years is not complete. This data is not adequate for the construction of networks, because the gaps in the citation collection would cause artificial cut-off points in the networks. Also, it was collected over a range of social science disciplines, so that data for any one discipline is thin. Data for the main citation study was collected for 1969-1970 only; this data by itself is again not suitable for building networks, since the data must be available over a number of years. DISISS will therefore investigate the possibility of obtaining data from other sources, perhaps ISI, for construction of citation networks.

REFERENCES

BROWN, C.H. Scientific serials: characteristics and lists of most cited publications in mathematics, physics, chemistry, geology, physiology, botany, zoology, and entomology. Chicago, Association of College and Reference libraries, 1956.

EAST, H. & WEYMAN, A. A study of the source literature of plasma physics. Aslib Proceedings, 1969, 21, 160-71.

GARFIELD, E. Citation indexes in sociological and historical research. American Documentation, 1963, 14(4), 289-290.

GARFIELD, E. Citation indexes in sociological and historical research. Current Contents/Life Sciences, 1969, 12 (34), p. 4.

GARFIELD, E. Citation indexing, historio-bibliography, and the sociology of science. Proceedings of the Third International Congress of Medical Librarianship. Amsterdam, 5-9 May 1969. K. E. Davis and W. D. Sweeney (eds.) Amsterdam, Excerpta Medica, 1970, 187-204.

GARFIELD, E. The Science Citation Index as a Blueprint of the journal literature. Institute of Scientific Information, 1971.

GARFIELD, E. "Explosive-welding" search reverberates in Moscow seminar on citation networks. Current Contents, 1972, (4(35), 6-10.

GARFIELD, E., and SHER, I.H. New factors in the evaluation of scientific literature through citation indexing. American Documentation, 1963, 14, 195-201.

GARFIELD, E., I. H. SHER & R. J. TORPIE. The use of citation data in writing the history of science. Philadelphia, Institute for Scientific Information, 1964.

GARFIELD, E. et al. Repartee. Journal of Library History, 1970, 5(2), 184-188.

HALL, A.M. The use and value of citations: a state of the art report. London, The Institution of Electrical Engineers, 1970.

(References cont'd..)

KESSLER, M.M. Bibliographic coupling between scientific papers. American Documentation, 1963, 14, 10-25.

MARTYN, J. An examination of citation indexes. Aslib Proceedings, 1965, 17(6) 184-196.

NARIN, F. CARPENTER, M., BERLT, N.C. Interrelationships of scientific journals. Journal of the American Society for Information Science, 1972, 23(5), 323-331.

PRICE, D. J. de SOLLA. Networks of scientific papers. Science, 1965, 149(3683) 510-515.

PRICE, N. and SCHIMINOVITCH, S. A clustering experiment: first step towards a computer generated classification. Information storage and retrieval, 1968, 4, 271-280.

SPENCER, C.C. Subject searching with SCI. Preparation of a drug bibliography using Chemical Abstracts, Index Medicus and Science Citation Index 1961-4. American Documentation, 1967, 18(2), 87-96.

WALDHART, T. J. A preliminary analysis of Science Citation Index. Research paper in Library Science, University of Wisconsin, 1964.

APPENDIX A.

REFERENCES - SHORT TERM MEMORY CITATION NETWORK

1. PETERSON, L. R. & PETERSON, M.J. Short-term retention of individual verbal items. Journal of Experimental Psychology, 1959, 58, 193-198.

2. LLOYD, K. E., REID, L. S. & FEALLOCK, J.B. Short-term retention as a function of the average number of items presented. Journal of Experimental Psychology, 1960, 60, 201-207.

3. MURDOCK, B.B., Jr. Short-term retention of single paired associates. Psychological Reports, 1961, 8, 280.

4. KEPPEL, G. & UNDERWOOD, B. J. Proactive inhibition in short-term retention of single verbal items. Journal of Verbal Learning and Verbal Behaviour, 1962, 1, 153-161.

5. MERTON, A. W. Implications of short-term memory for a general theory of memory. Journal of Verbal Learning and Verbal Behaviour, 1963, 2, 1-21.

6. MURDOCK, B.B., Jr. Short-term retention of single paired associates. Journal of Experimental Psychology, 1963, 65, 433-443.

7. MURDOCK, B.B., Jr. Short-term memory and paired associate learning. Journal of Verbal Learning and Verbal Behaviour, 1963, 2, 320-328.

8. SCHUMSKY, D.A., GRASHA, D.G. & ALLEN, C.K. Proactive inhibition and item similarity in short-term memory. Journal of Verbal Learning and Verbal Behaviour, 1963, 2, 440-445.

(Appendix A cont'd..)

9. WICKENS, D.D. BORN, D.G. and ALLEN, C.K. Proactive inhibition and item similarity in short-term memory. Journal of Verbal Learning and Verbal Behaviour, 1963, 2, 440-445.

10. CONRAD, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

11. LOESS, H. Proactive inhibition in short-term memory. Journal of Verbal Learning and Verbal Behaviour, 1964, 3, 362-368.

12. MURDOCK, B.B., Jr. Proactive inhibition in short-term memory. Journal of Experimental Psychology, 1964, 68, 184-189.

13. POSTMAN, L. Short-term memory and incidental learning. In A. Melton (ed.), Categories of human learning. New York, Academic Press 1964.

14. HOWE, M.J.A. Intra-list differences in short-term memory. Quarterly Journal of Experimental Psychology, 1965, 17, 338-342.

15. KEPPEL, G. Problems of method in the study of short-term memory. Psychological Bulletin, 1965, 63, 1-13.

16. MURDOCK, B.B., Jr. Effects of a subsidiary task on short-term memory. British Journal of Psychology, 1965, 56, 413-419.

17. POSNER, M.I. & ROSSMAN, E. Effect of size and location of informational transforms upon short-term retention. Journal of Experimental Psychology, 1965, 70, 496-505.

18. ROHRMEN, N.L. & JAHNKE, J.C. Effect of recall condition, presentation rate and retention interval on short-term memory. Psychological Reports, 1965, 16, 877-883.

19. WICKELGREN, W.A. Acoustic similarity and retroactive interference in short-term memory. Journal of Verbal Learning and Verbal Behaviour, 1965,

20. CRAWFORD, J., HUNT, E. & PEAK, G. Inverse forgetting in short-term memory. Journal of Experimental Psychology, 1966, 72, 415-422.

21. GOGGIN, J. Retroactive and proactive inhibition in short-term retention of paired associates. Journal of Verbal Learning and Verbal Behaviour, 1966, 5, 526-534.

22. PETERSON, L.R. Short-term verbal memory and learning. Psychological Review, 1966, 73, 193-207.

23. POSNER, M.I. & KONICK, A.F. On the role of interference in short-term retention. Journal of Experimental Psychology, 1966, 72,221-231.

24. WICKELGREN, W.A. Phonemic similarity and interference in short-term memory for single letters. Journal of Experimental Psychology, 1966, 71, 396-404.

25. AARSONSON, D. Temporal factors in perception and short-term memory. Psychological Bulletin, 1967, 130-144.

26. BJORK, R.A. The effect of instructions to selectively forget during short-term memory. Memorandum Report No. 3, 1967, University of Michigan, Human Performance Centre.

27. LOESS, H. Short-term memory, word class, and sequence of items. Journal of Experimental Psychology, 1967, 74, 556-561.

28. LOESS, H., and WAUGH, N.C. Short-term memory and intertrial interval. Journal of Verbal Learning and Verbal Behaviour, 1967, 6, 544.

29. TALLAND, G.A. Short-term memory with interpolated activity. Journal of Verbal Learning and Verbal Behaviour, 1967, 6, 144-150.

30. BJORK, R.A., LABERGE, D.& LEGRAND, R. The modification of short-term memory through instructions to forget. Psychonomic Science, 1968, 10, 55-56.

(Appendix A. cont'd..)

31. JAHNKE, J.C. Delayed recall and the serial-position effect of short-term memory. Journal of Experimental Psychology, 1968, 76, 618-622.

32. MARTIN, E., ROBERTS, K.H.& COLLINS, A.M. Short-term memory for sentences. Journal of Verbal Learning and Verbal Behaviour, 1968, 7, 560-566.

33. MURDOCK, B.B., Jr. Serial order effects in short-term memory. Journal of Experimental Psychology, 1968, 76 (4 pt.2)

34. WICKENS, D.D., and CLARKE, S.F. Osgood dimensions as an encoding class in short-term memory. Journal of Experimental Psychology, 1968, 78, 580-584.

35. DICKENS, D.W., CLARK, S.E., HILL, F.A. and WILLINGER, R.P. An investigation of grammatical class as an encoding class in short-term memory. Journal of Experimental Psychology, 1968, 78, 599-604.

36. BADDELEY, A.D., SCOTT, D.& DRYNAN, R. Short-term memory and limited capacity hypothesis. British Journal of Psychology, 1969, 60, p.51.

37. BLOUNT, W.R. Short-term memory in retardates as a function of direction and delay of recall. Psychonomic Science, 1969, 14, p.69.

38. CERMAK, L.S. Repetition and encoding in short-term memory. Journal of Abnormal Psychology, 1969, 92, p.321.

39. DILLON, R.F. & REID, L.S. Short-term memory as a function of information processing during retention interval. Journal of Experimental Psychology, 1969, 91. p.261.

(Appendix A. cont'd...)

40. DOLL, T.J. Short-term retention - preparatory set as covert rehearsal Journal of Experimental Psychology, 1969, 82, p.175.

41. ELLIS, N.R. Effects of interpolated recall on short-term memory. Journal of Experimental Psychology, 1969, 79, p.568.

42. ELMES, D.G. Role or prior calls and storage load in short-term memory. Journal of Experimental Psychology, 1969, 79, p.468.

43. GILSON, E.Q. Actile short-term memory. Quarterly Journal of Experimental Psychology, 1969, 21, p.180.

44. LACHAR, B. & GOGGIN, J. Effects of changes in word length on proactive interference in short-term memory. Psychonomic Science, 1969, 17, p.213.

45. LANDAUER, T.K. Relation between short and long-term memory - effect of order on retroactive interference. Psychonomic Science, 1969, 17, p.116.

46. MARTIN, E. & WALTER, D.A. Subject uncertainty and word-class effects in short-term memory for sentences. Journal of Experimental Psychology, 1969, 80, p.47.

47. NELSON, T.O. & BATCHELDER, W.H. Forgetting in short-term recall - all or none on decremental. Journal of Experimental Psychology, 1969, 82, p.96.

48. RYAN, J. Temporal grouping rehearsal and short-term memory. Quarterly Journal of Experimental Psychology, 1969, 21, p.148.

(Appendix A. cont'd...)

49. RYAN, J.F., Short-term memory and rehearsal in educable subnormals. American Journal of Mental Deficiency, 1969, 74, p.218.

50. SCHUMSKY, D.A., GRASHA, A.F., TRINDER, J., RICHMAN, C. List length and single-trial short-term memory. Journal of Experimental Psychology, 1969, 82, p.238.

51. SCHWERTZ, F. Some Problems and notes about short-term memory. Psychological Reports, 1969, 24, p.71.

52. STANNERS, R.F. & MEUNIER, G.F. Pronunciability and rehearsal time in short-term memory with controlled acquisition. Journal of Experimental Psychology, 1969, 80, p.359.

53. STANNERS, R.F., MEUNIER, G.F. & HEADLEY, D.B. Reaction time as an index of rehearsal in short-term memory. Journal of Experimental Psychology, 1969, 82, p.566.

54.. TARPY, R.M. Motivation and short-term retention - evidence for covert rehearsal. American Journal of Psychology, 1969, 82, p.111.

55.. TARPY, R.M. Effects of interference in STM with minimum covert rehearsal. Psychological Reports, 1969, 24, p.723.

56.. TURVEY, M.T. Connotative classification and proactive interference in short-term memory. Psychonomic Science, 1969, 16, p.223.

57. TURVEY, M.T. & WITTLINGER, R.D. Attenuation of proactive interference in short-term memory as a function of cueing to forget. Journal of Experimental Psychology, 1969, 80, p.295.

APPENDIX B.

REFERENCES - DEVIANT BEHAVIOUR CITATION NETWORK.

1. ERIKSON, K.T. Notes on the sociology of deviance. Social Problems, 1962, 9, 307-14.

2. KITSUSE, J.I. Societal reactions to deviant behaviour : problems of theory and method. Social Problems, 1962, 9, 247-56.

3. CLINARD, M.B. .ed. Sociology of deviant behaviour. New York: Holt, Rinehart & Winston, 1963.

4. BECKER, H.S. Outsiders: studies in the sociology of deviance. New York: The Free Press of Glencoe, 1963.

5. CLINARD, M.B. (ed.) Anomie and deviant behavior: a discussion and critique. New York: The Free Press of Glencoe, 1964.

6. COHEN, A.K. The sociology of the deviant act: anomie theory and beyond. American Sociological Review, 1965, 30, 5-14.

7. WILKINS, L.T. Social deviance: social policy, action and research. Englewood Cliffs, N.J.: Prentice-Hall, 1965.

8. SCHUR, E.M. Crimes without victims : deviant behavior and public policy. Englewood Cliffs, N.J. : Prentice Hall, 1965.

9. COHEN, A.K. Deviance and control. Englewood Cliffs, N.J.; Prentice-Hall, 1966.

10. GIBBS, J.P. Conceptions of deviant behavior : the old and the new. Pacific Sociological Review, 1966, 9, 9-14.

11. LEMERT, E.M. Human deviance, social problems and social control. Englewood Cliffs, N.J.; Prentice-Hall, 1967.

(Appendix B. cont'd..)

12. RUBINGTON, E. & WEINBERG, M.S. (eds.) Deviance : the interactionist perspective. New York : Macmillan, 1968.

13. SCHUR, E.M. Reactions to deviance : a critical assessment. American Journal of Sociology, 1969-70, 309-322.

14. SHAH, S.A. Crime and mental illness - some problems in defining and labelling deviant behaviour. Mental Hygiene, 1969, 53, p.21.

15. WEISER, S. Deviance and informal sanctions in mental illness. Bibliotheca Psychiatrica et Neurologica, 1969, p.24.

APPENDIX C.

List of DISISS Working Papers

No.1. Comparison of results of science user studies with "Investigation into Information Requirements of the Social Sciences". July, 1971.

No.2. A machine readable data base of social science serials. November, 1971.

No.3. The relationship between primary and secondary literature in the social sciences: a study of secondary literature in criminology. June, 1972.

No.4. Characteristics of citations in social science monographs. June, 1972.

No.5. Citation patterns in the social sciences: results of pilot citation study and selection of source journals for main citation study. October, 1972.

No.6. The use of citation linkages and networks for information retrieval in the social sciences, March 1973.

No.7. Size of monograph literature in social science (in preparation).

No.8. CLOSSS: a machine readable data base of social science serials: progress report 1971-1972. April 1973.

No.9. Monograph titles cited in social science journal articles.

No.10. The relationship of frequency of citation to use and value (in preparation).